HELDA
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Glycaemic variability-based classification of impaired glucose tolerance vs. type 2 diabetes using continuous glucose monitoring data

Longato, Enrico

2018-05-01

Longato , E , Acciaroli , G , Facchinetti , A , Hakaste , L , Tuomi , T , Maran , A & Sparacino , G 2018 , ' Glycaemic variability-based classification of impaired glucose tolerance vs. type 2 diabetes using continuous glucose monitoring data ' , Computers in Biology and Medicine , vol. 96 , pp. 141-146 . https://doi.org/10.1016/j.compbiomed.2018.03.007

http://hdl.handle.net/10138/302391
https://doi.org/10.1016/j.compbiomed.2018.03.007

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Computers in Biology and Medicine 96 (2018) 141–146

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Glycaemic variability-based classification of impaired glucose tolerance vs. type 2 diabetes using continuous glucose monitoring data

Enrico Longato [a], Giada Acciaroli [a], Andrea Facchinetti [a], Liisa Hakaste [b,c], Tiinamaija Tuomi [b,c,d], Alberto Maran [e], Giovanni Sparacino [a,*]

[a] Department of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131, Padova, Italy
[b] Endocrinology, Abdominal Centre, University of Helsinki and Helsinki University Hospital, Haartmaninkatu 8, FI-00014, Helsinki, Finland
[c] Folkhälsan Research Center and Research Program for Diabetes and Obesity, University of Helsinki, Haartmaninkatu 8, FI-00014, Helsinki, Finland
[d] Finnish Institute for Molecular Medicine, University of Helsinki, Tukholmankatu 8, FI-00014, Helsinki, Finland
[e] Department of Medicine, University of Padova, Via Giustiniani 2, 35128, Padova, Italy

ARTICLE INFO

Keywords:
Classification
Continuous glucose monitoring
Diabetes
Glycaemic variability
Support vector machine

ABSTRACT

Many glycaemic variability (GV) indices extracted from continuous glucose monitoring systems data have been proposed for the characterisation of various aspects of glucose concentration profile dynamics in both healthy and non-healthy individuals. However, the inter-index correlations have made it difficult to reach a consensus regarding the best applications or a subset of indices for clinical scenarios, such as distinguishing subjects according to diabetes progression stage. Recently, a logistic regression-based method was used to address the basic problem of differentiating between healthy subjects and those affected by impaired glucose tolerance (IGT) or type 2 diabetes (T2D) in a pool of 25 GV-based indices. Whereas healthy subjects were classified accurately, the distinction between patients with IGT and T2D remained critical. In the present work, by using a dataset of CGM time-series collected in 62 subjects, we developed a polynomial-kernel support vector machine-based approach and demonstrated the ability to distinguish between subjects affected by IGT and T2D based on a pool of 37 GV indices complemented by four basic parameters—age, sex, BMI, and waist circumference—with an accuracy of 87.1%.

1. Introduction

So-called continuous glucose monitoring (CGM) systems have enabled the quasi-continuous (up to a resolution of a sample per minute), real-time monitoring of blood glucose (BG) concentrations [1]. Currently, the majority of available commercial products comprise three main elements: a minimally-invasive subcutaneous electrochemical needle sensor (to be inserted into the abdomen or arm), wireless transmitter, and receiver, which is often coupled with a display or other user interface [2,3]. These existing devices have accuracies similar to those of standard finger-prick BG devices (see Refs. [4–7] for example references).

The concept of glycaemic variability (GV) as a tool for characterising the dynamic properties of BG concentration traces has gained importance in light of the increased use and reliability of CGM systems [8,9]. Previous reports have proposed tens of GV indices with varying degrees of mathematical complexity and clinical interpretability (e.g. [10], and [11]), and some are thought to serve as markers of a number of adverse pathological outcomes and complications in diabetic patients [12–16]. However, a general consensus regarding the best subset of GV indices and the exact combinations needed to evaluate the metabolic state of a subject has not yet been reached [17,18], partly because of the high degrees of correlation between many GV metrics [10,19,20].

To the best of our knowledge, Ref. [21] presents the first investigation of a basic classification problem, such as distinguishing between healthy subjects and patients affected by impaired glucose tolerance (IGT) or type 2 diabetes (T2D), using a large subset of GV indices. In that work, the authors implemented a cascade of two logistic regression steps to sequentially determine which of the 102 subjects were healthy and subsequently pinpoint the exact altered metabolic state (IGT or T2D) of the remaining subjects. The results achieved in that study were promising in terms of the highly accurate identification of healthy subjects (91.4%

* Corresponding author.
E-mail addresses: enrico.longato.1@phd.unipd.it (E. Longato), giada.acciaroli@phd.unipd.it (G. Acciaroli), facchine@dei.unipd.it (A. Facchinetti), liisa.hakaste@hus.fi (L. Hakaste), tiinamaija.tuomi@hus.fi (T. Tuomi), alberto.maran@unipd.it (A. Maran), gianni@dei.unipd.it (G. Sparacino).

https://doi.org/10.1016/j.compbiomed.2018.03.007
Received 23 October 2017; Received in revised form 10 March 2018; Accepted 10 March 2018

0010-4825/© 2018 Elsevier Ltd. All rights reserved.

cross-validation accuracy); however, the distinction between subjects affected by IGT and T2D proved quite critical (79.5% ± 15.9% cross-validation accuracy).

To assess the reasons underlying the difficult CGM-based distinction between IGT and T2D, let us consider the traces in Fig. 1. The top panel shows a CGM trace recorded in a representative subject affected by T2D who was frequently affected by hyperglycaemia (BG > 180 mg/dL), which is characteristic of the condition. The middle panel shows a CGM trace recorded from a patient affected by IGT; here, we can appreciate the somewhat expected reduction in the frequency and entity of hyperglycaemic episodes, although a brief hypoglycaemic event (BG < 70 mg/dL) occurred. Finally, the bottom panel features a CGM trace which is arguably much more similar to the trace acquired from the patient with IGT, despite actually having been acquired from a subject with full-fledged T2D. Clearly, the distinction of IGT and T2D based on an inspection of CGM signals is a non-trivial problem attributable to subtle differences that may not be readily detectable, but nonetheless must be assessed quantitatively in terms of GV indices.

In the present paper, which considers a dataset of 62 IGT and T2D subjects evaluated previously by Acciaroli et al. [21], we discuss the use of a machine learning approach to further develop the previous work, with the aim of improving the distinction between IGT and T2D. Specifically, we explored two possible avenues of improvement: first, we attempted to increase the quantity of information available for classification by considering a wider subset of GV indices in the literature and complementing these tools using basic information such as age, sex, body mass index (BMI), and waist circumference; second, we tested the performance of more sophisticated techniques— namely, two support vector machine (SVM) variants—which might be more suitable for more challenging classification problems. We demonstrate that IGT and T2D patients could be distinguished with accuracy levels as high as 87.1%; this improvement was significant when compared to the baseline of 61.3% established by implementing a logistic regression-based strategy similar to that proposed in Ref. [21].

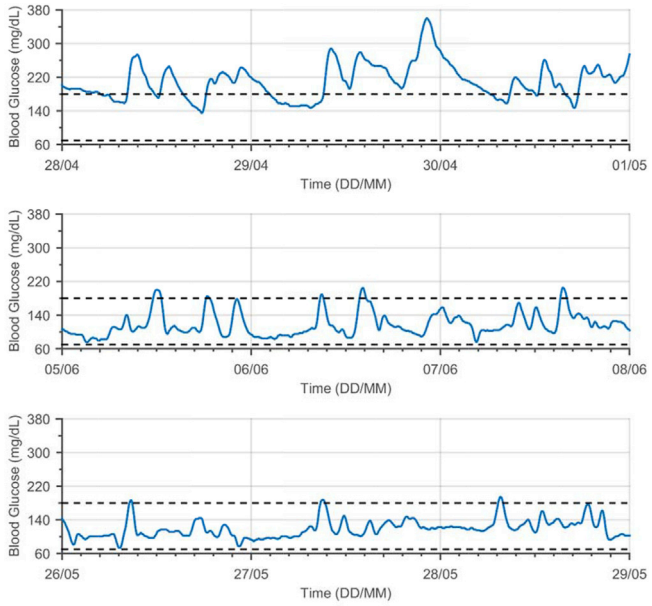

Fig. 1. Three exemplificative continuous glucose monitoring (CGM) traces. The top and bottom panels show CGM traces acquired from two subjects affected from T2D; the middle panel shows a CGM trace from a patient with IGT. The dashed lines represent the hyperglycaemia (blood glucose >180 mg/dL) and hypoglycaemia (blood glucose <70 mg/dL) thresholds.

2. Database

2.1. Dataset description

The 62 subjects analysed in the present work represent a subpopulation extracted from two long-term studies, the Botnia Prospective Study (BPS) and the Botnia PPP Study (approved by the ethics committee of Helsinki University Hospital; all study participants provided informed consent). The data, which were acquired during the EU FP7 Mosaic Project [22], include both CGM traces and a wide range of clinical parameters (including age, sex, BMI, and waist circumference) obtained during two annual visits per subject (reference timeframes: February–June 2014 and March–June 2015). All subjects were constantly monitored using the iPro CGM system (Medtronic MiniMed, Inc., Northridge, CA, USA) for 6-days periods at a frequency of one sample every 5 min. IGT or T2D were diagnosed at the time of the visit using an oral glucose tolerance test (OGTT) [23]. At the time of the first visit, 36 and 26 subjects were affected by IGT and T2D; 1 year later, two and one subject experienced diagnostic changes from T2D to IGT and from IGT to T2D, respectively, yielding a total of 37 IGT and 25 T2D patients.

2.2. Glycaemic variability indices

CGM-acquired signals were initially processed to extract a set of 37 GV indices (including the 25 indices used previously used in Ref. [21] and, for purposes of compressed representation, in Refs. [19,20]).

Seven of those indices comprise a subset that is related, roughly speaking, to the 'gross' statistical properties of CGM signals: mean, standard deviation (SD), coefficient of variation (CV), median, range, interquartile range (IQR) [19,24], and J-index [25]. We also considered seven metrics related to variations in the times of global properties, point-values, and percentages of time spent in the hypo-, eu-, and hyper-glycaemic ranges: SD_w and SD_d [26], CONGA and MODD [27], and $\%BG_{below}$, $\%BG_{above}$, and $\%BG_{within}$ [19,28]. A third subset of four GV indices addresses the identification of the entities and frequencies of peak-nadir excursions: the MAGE, MAGE+ and MAGE- [29,30] and excursion frequency (EF) [31–33] indices. Twelve additional indices are derived from empirical transformations of the glycaemic scale intended to emphasise the abnormality of hypoglycaemia and hyperglycaemia relative to the physiological state of euglycaemia: M-value [34], GRADE, $GRADE_{hypo}$, $GRADE_{hyper}$, $GRADE_{eu}$ [35], Hypo Index, Hyper Index, IGC [36], LBGI, HBGI, BGRI, and ADRR [37,38]. Finally, we computed seven indices as the moment invariants (φ_1, …, φ_7) proposed in Ref. [39] for pattern recognition; these indices provide a stable description of the binary image, obtained by assigning the values 1 and 0 to all pixels below and above the CGM trace, respectively, in line with the rationale followed in Refs. [31–33].

3. Methods

3.1. Classification strategy

We defined training and test sets comprising data acquired during the first and second visits, respectively, yielding a total of 62 samples per set. We decided against mixing data from different visits to simulate a dataset that would be realistically available and applicable to clinical practice (i.e., no clinician should be expected to conduct a longitudinal trial to acquire a consistent training set). Different subsets of our total set of 37 GV indices (see Section 4 for an in-depth discussion of subset composition) and four basic parameters (age, sex, BMI, and waist circumference) were used as features, whereas the results of the OGTT, a reference diagnostic technique for both IGT and T2D [23], were used as the label (ground truth).

Throughout our analyses, we developed and tested several variations of three popular classification techniques: logistic regression, a soft-margin SVM, and a SVM to which the kernel trick was applied using a

polynomial kernel [40]. To prevent overfitting, we used L2 regularisation by estimating the penalty coefficient C together with other hyper-parameters, if present.

Specifically, using x_i for i = 1, …, 62, y_i as the feature vector associated with the i-th subject in the training set, and C as the L2 regularisation coefficient, the following three cost functions were applied.

For logistic regression, we minimised

$$L = \sum_{i=1}^{62}[y_i \log(\sigma(w^T x)) + (1 - y_i)\log(1 - \sigma(w^T x))] + \frac{1}{2}C||w||^2 \quad (1)$$

where w is the weight vector, which also includes the intercept parameter, and σ(·) is the sigmoid function.

For the soft-margin SVM, we maximised

$$L = \sum_{i=1}^{62}\alpha_i - \frac{1}{2}\sum_{i=1}^{62}\sum_{j=1}^{62}\alpha_i\alpha_j y_i y_j x_i^T x_j \quad s.t. \begin{cases} \sum_{i=1}^{62}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \end{cases} \quad (2)$$

where α_i for i = 1, …, 62 are the sparse dual-problem weights equivalent to w.

For the SVM with the kernel trick formulation, we maximised

$$L = \sum_{i=1}^{62}\alpha_i - \frac{1}{2}\sum_{i=1}^{62}\sum_{j=1}^{62}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \quad s.t. \begin{cases} \sum_{i=1}^{62}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \end{cases} \quad (3)$$

where we replaced the inner product, $x_i^T x_j$, with the polynomial kernel

$$k(x_i, x_j) = (x_i^T x_j + c_0)^d \quad (4)$$

We used a grid search to estimate the hyper-parameters; specifically, we tested the performances of different combinations of these values in a four-fold stratified cross-validation setting. In practice, we first divided the training set into four folds while preserving the original ratio of IGT to T2D subjects within each fold; next, we fixed the value of the hyper-parameters to some point in the grid, estimated four weight vectors, w, by training the classifier on three of four folds and testing it on the fourth fold, and calculated the average inter-fold error; finally, after repeating this procedure for each point in the hyper-parameter grid, we selected the set of hyper-parameters with which we obtained the minimum average cross-validation error. The final value of the weight vector, w, was eventually selected by minimising the cost function using the final set of hyper-parameters. By applying this rather complex procedure and alternating between the cost function and average cross-validation error minimisation, we attempted to maximise the generalisation potential of our models to ensure efficacy despite the limited cardinality of the available training set.

3.2. Performance assessment

After estimating the weight vector and appropriate set of hyper-parameters for each model, we conducted a performance assessment that considered two distinct metrics to gain insights regarding the goodness of classification, particularly test accuracy and cross-validation accuracy.

Here, test accuracy simply refers to classification accuracy, defined as the percentage of correctly classified subjects over the training set:

$$ACC = \frac{1}{N}\sum_{i=1}^{N} I(\hat{y}_i = y_i) \quad (5)$$

where I(·) is the binary indicator function, which is equal to 1 when the condition is true (i.e., when the estimated label $\hat{y}_i$ is correct), and N is the cardinality of the test set (in this case, N = 62). As ACC is assessed using an actual test set (i.e., a hitherto unseen collection of samples), it is indicative of the real-life classification performance.

We computed both the mean and inter-fold standard deviation of cross-validation accuracy using a four-fold stratified cross-validation. Specifically, we first divided the entire training set in four folds while maintaining a consistent ratio of IGT to T2D examples in each fold. Here, k = 1, …, 4 is the index associated with each fold, and N_k is the number of elements in the k-th fold. To compute the cross-validation accuracy for each model, we iteratively set aside the k-th fold of our training set to be used in lieu of an actual test set, re-trained the model using the remaining three folds, and computed the resulting accuracy, ACC_k, for k = 1, …, 4 by applying Equation (5) with N = N_k. Cross-validation accuracy was then defined as

$$CVACC = \frac{1}{4}\sum_{k=1}^{4} ACC_k \quad (6)$$

Additionally, we computed the inter-fold standard deviation of CVACC as

$$std(CVACC) = \sqrt{\frac{1}{4}\sum_{k=1}^{4}(ACC_k - CVACC)^2} \quad (7)$$

We note that although the technique used to assess CVACC was identical to the technique described during the grid-search step of the model estimation phase, these assessments were undertaken independently. In other words, the complete calculation of CVACC requires two nested cross-validation loops, the innermost of which accesses only three quarters of the total training set. Given the further reduction in cardinality of the training portion of the data, we would expect CVACC to be pessimistic relative to the test set accuracy ACC. However, the standard deviation of CVACC is a good measure of the expected variability in our test set point estimate.

4. Results

4.1. Baseline GV-based classification

To establish a reference, we first re-implemented a logistic regression-based strategy using the same subset of 25 indices considered in Ref. [21]. As expected, the ACC results were unsatisfactory; here, a simple logistic regression with an ACC of 61.3% barely out-performed a random classifier when confronted with an independent test set. We note that some level of discrepancy emerged between ACC and CVACC; however, as the test set accuracy ACC is representative of a real-life use-case (i.e., the application of a trained classifier to an unknown dataset), we considered it to be the more reliable point-value metric. Additionally, we referred to CVACC to check for excessive variability or otherwise extremely unsatisfactory results.

Given this preliminary set of results, we could argue that a simple linear classification technique, such as logistic regression, might be insufficient to capture the subtle differences in GV that are needed to accurately differentiate between IGT and T2D (recall Fig. 1). Based on these considerations, we attempted to overcome this limitation by adopting a more complex model, namely a SVM with a polynomial kernel, of which the main advantage (relative to logistic regression) is the ability to identify nonlinear boundaries between classes [38]. However, the application of the polynomial-kernel SVM to the same subset of 25 GV indices yielded inconclusive results; specifically, a noticeable increase in CVACC (from 71.1% to 79.2%) was offset by the quasi-random test set performance (CVACC = 56.5%).

4.2. Extended pool of GV indices

To investigate the reasons underlying this apparently inconsistent behaviour, we attempted to enrich the information available to the polynomial-kernel SVM by increasing the number of considered GV

Table 1
Summary of the results. The first and second columns report the model names and feature sets, respectively, while the third and fourth columns report cross-validation and test set accuracy, respectively.

Classifier	Features (#)	CVACC	ACC
Logistic Regression	GV indices (25)	71.1 (±6.4) %	61.3%
SVM (poly. kernel)	GV indices (25)	79.2 (±4.8) %	56.5%
SVM (poly. kernel)	GV indices (37)	71.1 (±8.9) %	71.0%
SVM (poly. kernel)	GV indices (37) + parameters (4)	72.5 (±7.5) %	87.1%

indices to 37. We hypothesised that a more complex classifier could successfully integrate a wider selection of features, represent a more nuanced picture of the examined CGM traces, and achieve more stable and satisfactory results. As expected, the development and application of our 37-index polynomial-kernel SVM allowed us to reconcile CVACC and ACC, which were both approximately equal to 71%.

All the relevant results discussed so far are reported in Table 1. A survey of this table demonstrates how the 37-index polynomial-kernel SVM clearly outperforms both the logistic regression baseline and the initially devised 25-index polynomial-kernel SVM extension. This suggests that a wider subset of GV indices is instrumental to a better characterisation of the subtle differences between the CGM traces acquired from subjects affected by IGT and T2D, especially when these data are fed into a more complex classifier. However, the performance remained far from satisfactory, which may indicate that GV alone is not sufficient to completely characterise a subject as affected by IGT or T2D. In the following section, we will attempt to improve the classification performance by considering additional GV-independent variables.

4.3. Inclusion of additional parameters: age, sex, BMI, and waist circumference

Previous studies [41,42] suggested the importance of considering subsets of GV indices to define glycated haemoglobin (HbA1c)-independent dimensions, which could help to comprehensively characterise the metabolic state of a subject. Here, we followed the same rationale and enriched our index pool with four basic parameters pertaining to each patient which are normally available without any cost: age, sex, BMI, and waist circumference. Thus, we aimed to complement the information provided by GV with basic clinical information regarding the subjects.

As described previously, we developed and tested a polynomial-kernel SVM using the whole set of 41 variables (37 GV indices plus the four basic parameters). Whereas CVACC differed only slightly when compared to the 37-index GV-only scenario (72.5% vs. 71.1%), we achieved substantial improvements in terms of ACC; specifically, 87.1% of the patients in the actual test set were correctly identified as being affected by IGT or T2D. A visual representation of the improvement, obtained via principal component analysis (PCA), is reported in Fig. 2, where we can qualitatively appreciate the increase in the discriminative power of our method by comparing the numbers of circles (hits) between the two panels. Furthermore, we highlight the fact that our SVM correctly labelled the CGM traces reported in Fig. 1, as opposed to the baseline logistic regression model discussed in Section 4.1 which mistakenly classified the second subject as affected by T2D.

Finally, we report that the inclusion of the four basic parameters introduced in this section to the logistic regression model, similar to the methodology in Ref. [21], did not yield satisfactory classification performance results. Indeed, ACC remained less than 68%, regardless of the initial pool of GV indices (25 as in Ref. [21], or 37 as investigated here).

5. Discussion

In the present work, we demonstrated the value of the concept of GV in highlighting subtle differences between the CGM traces acquired in subjects affected by IGT and T2D.

To the best of our knowledge, this particular problem was previously only tackled in Ref. [21], where Acciaroli et al. proposed a strategy based on a cascade of two logistic regression steps to first isolate healthy controls, and only then differentiate between patients with IGT and T2D. This latter distinction proved quite critical in terms of performance metrics, which was likely due to a combination of factors. The first factor is the high degree of correlation among GV indices, which was addressed using a descriptive rather than discriminative framework in Ref. [20], wherein the application of a sparse PCA to a pool of GV indices (quantified from CGM in 13 subjects affected by T2D) revealed that most of the variance could be explained by a subset of indices. The performance of the classifier proposed in Ref. [21] might also have been affected by the absence of linear (in terms of GV) separability between the classes defined by a diagnosis of IGT or T2D. This factor would also partially explain why the first logistic regression step in Ref. [21] achieved

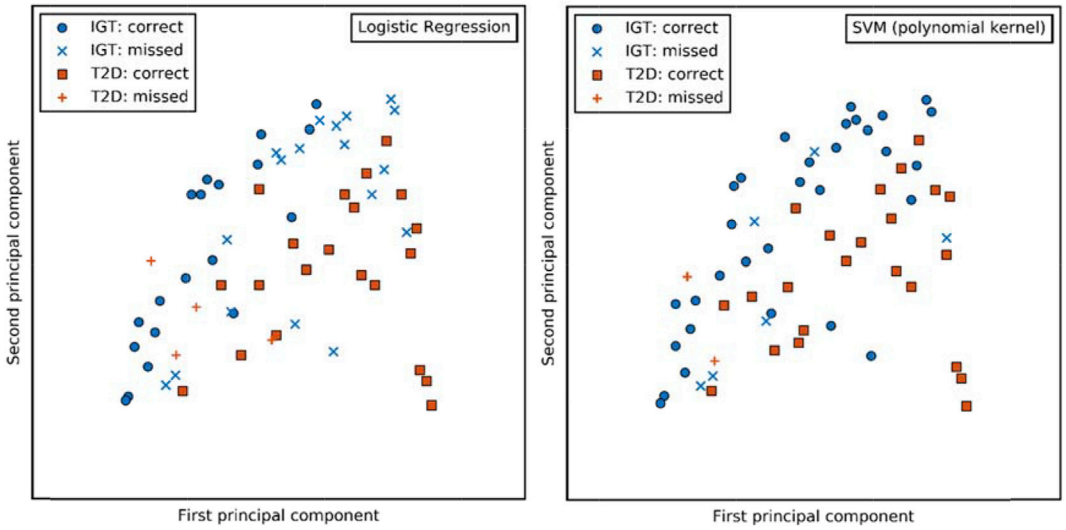

Fig. 2. Visualisation of the study results. We compared the baseline results obtained by applying a logistic regression to 25 GV indices (left panel) and our best performing model (right panel), a polynomial-kernel SVM, to 37 GV indices and four basic parameters. The circles and squares correspond respectively to correctly classified IGT and T2D subjects; crosses and plus signs indicate respective misclassifications. The two-dimensional projection was obtained using a PCA of the test set, to which the slightly different shapes of the two point clouds are attributed.

drastically better results than the second one. In all likelihood, the distinction between healthy and non-healthy subjects is actually a matter of defining a threshold hyperplane in the feature spaces of GV indices, whereas the border separating IGT and T2D is probably a more complex n-dimensional surface.

The non-linear models applied to GV indices were also investigated in Refs. [31,32], which yielded very promising results in the context of characterising the quality of glycaemic control in patients affected by T1D or T2D. Although this problem differs fundamentally from the problem investigated in the present work, both required researchers to address the blurry boundaries between classes. In addition to the intrinsic difficulties in quantifying glycaemic control, however, Marling et al. also reported an average intra-clinician consistency of 82% [32]; in other words, the same expert, when tasked with classifying the same CGM trace, often reconsidered his or her own assessment. In light of this finding, the distinction between IGT and T2D seems an appropriate task with which to evaluate the potential of GV indices to solve nuanced classification problems, while removing any noise that might be introduced by subjective labelling.

A possibly insightful by-product of a study on GV-based classification like that reported in the present paper could be the identification of an optimal subset of indices. To this end, in previous works several techniques, including sparse PCA [20], *t*-test filter [31], and backward elimination [31], were proposed, but none of them gave conclusive results. An alternative approach might be the assessment of relative feature importance via exhaustive search [40]. However, we did not implement it here, given the high number of covariates in the starting pool (41) which would have required evaluating about $2{\cdot}10^{12}$ different feature combinations.

Finally, although we realise that the sample size of our study may be a limiting factor in terms of generalising our results to a general population, we clarify that to the best of our knowledge, no broader data sets relevant to our scope are available even in the most recent literature. For instance, although [43] describes 63 subjects affected by T1D and T2D who were monitored by CGM, only HbA1c measurements were collected during follow-up visits. In Ref. [44], the CGM traces of 54 subjects affected by T2D and pre-diabetes were recorded, but no follow-up data were made available. Although a dataset with higher cardinality was recently used in Ref. [45], it comprised only subjects affected by T2D and did not include any follow-up data; therefore, this dataset was unsuitable for a classification task such as that considered in the present paper.

6. Conclusions

The distinction of subjects affected by IGT or T2D using CGM-extracted GV indices remains poorly investigated, and has been only partially addressed in Ref. [21], which yielded encouraging but not completely satisfactory results. In the present work, we aimed to improve upon those preliminary results by first abandoning logistic regression in favour of more sophisticated machine learning techniques and then introducing additional variables to complement the information carried by GV indices. After establishing a baseline using a similar methodology to that of [21], we achieved an initial improvement in performance by implementing a polynomial-kernel SVM on an extended subset of 37 GV indices, which increased the accuracy from 61.3% to 71.0%. The subsequent inclusion of four basic parameters–age, sex, BMI, and waist circumference–allowed us to solve the classification problem with a more satisfactory test set accuracy of 87.1%. Apparently, then, the subtle (in terms of CGM traces) boundaries between IGT and T2D are better defined in light of an overview of the subject's general metabolic state, which can be achieved using a wide set of variables comprising both GV indices and other patient-related information.

Future research will investigate the trade-off between performance and interpretability, which remains an open issue in that further investigation is needed to assess how accuracy would be affected by any model simplification in the context of each specific application. In fact, our inability to identify a stable optimal subset of features for this task was a direct result of using the kernel trick to boost the performance of our SVM model. Moreover, the limited number of samples at our disposal hindered the identification of the optimal strategy for the GV-based distinction of subjects affected by IGT or T2D. Accordingly, despite our promising results and satisfactory performance in the considered subpopulation, our methodology, or any extension of it, should be thoroughly validated before applying it to the general population, and adjustments should be made as appropriate. Additionally, a dataset with higher cardinality would be instrumental in the refinement of GV-only methods for IGT and T2D diagnoses that do not rely on clinical parameters other than CGM traces, especially if new GV metrics were developed in the process. Indeed, given the sheer number of indices in existence in the clinical literature and the variety of CGM signal properties they describe, a convincing case for a new indicator could only be made with substantial practical evidence in favour of its adoption. Hence, we refrained from proposing new metrics and, instead, stuck to analysing aspects of the CGM signal widely accepted as significant by the diabetes technology community.

Conflicts of interest

None Declared.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

We acknowledge the Botnia Study Group for recruiting and studying the participants of the Botnia Study (http://www.botnia-study.org/en/home/).

References

[1] E.R. Damiano, F.H. El-Khatib, H. Zheng, D.M. Nathan, S.J. Russell, A comparative effectiveness analysis of three continuous glucose monitors, Diabetes Care 36 (2) (Feb. 2013) 251–259.
[2] J.E. Lane, J.P. Shivers, H. Zisser, Continuous glucose monitors: current status and future developments, Curr. Opin. Endocrinol. Diabetes Obes. 20 (2) (Apr. 2013) 106–111.
[3] J. Kropff, J.H. DeVries, Continuous glucose monitoring, future products, and update on worldwide artificial pancreas projects, Diabetes Technol. Therapeut. 18 (S2) (Feb. 2016). S2–53-S2-63.
[4] M. Christiansen, et al., A new-generation continuous glucose monitoring system: improved accuracy and reliability compared with a previous-generation system, Diabetes Technol. Therapeut. 15 (10) (Oct. 2013) 881–888.
[5] G. Freckmann, S. Pleus, M. Link, E. Zschornack, H.-M. Klötzer, C. Haug, Performance evaluation of three continuous glucose monitoring systems: comparison of six sensors per subject in parallel, J. Diabetes Sci. Technol. 7 (4) (Jul. 2013) 842–853.
[6] T.S. Bailey, A. Chang, M. Christiansen, Clinical accuracy of a continuous glucose monitoring system with an advanced algorithm, J. Diabetes Sci. Technol. 9 (2) (Mar. 2015) 209–214.
[7] G. Cappon, G. Acciaroli, M. Vettoretti, A. Facchinetti, G. Sparacino, Wearable continuous glucose monitoring sensors: a revolution in diabetes treatment, Electronics 6 (3) (Sep. 2017) 65.
[8] J. Tay, C.H. Thompson, G.D. Brinkworth, Glycemic variability: assessing glycemia differently and the implications for dietary management of diabetes, Annu. Rev. Nutr. 35 (Jul. 2015) 389–424.
[9] M. DeGeeter, B.J. Williamson, L.M. Meadowcraft, A review of the current literature and guidance for treatment of the dawn phenomenon and glycemic variability, Diab. Obes. Metab. Disor. OA 2 (Nov. 2016) 50–59.
[10] D. Rodbard, The challenges of measuring glycemic variability, J. Diabetes Sci. Technol. 6 (3) (May 2012) 712–715.
[11] J.-P. Le Floch, L. Kessler, Glucose variability: comparison of different indices during continuous glucose monitoring in diabetic patients, J. Diabetes Sci. Technol. 10 (4) (Feb. 2016) 885–891.
[12] C. Gorst, et al., Long-term glycemic variability and risk of adverse outcomes: a systematic review and meta-analysis, Diabetes Care 38 (12) (Dec. 2015) 2354–2369.
[13] C.-R. Hsu, Y.-T. Chen, W.H.-H. Sheu, Glycemic variability and diabetes retinopathy: a missing link, J. Diabet. Complicat. 29 (2) (Mar. 2015) 302–306.

[14] J.E. Jun, et al., The association between glycemic variability and diabetic cardiovascular autonomic neuropathy in patients with type 2 diabetes, Cardiovasc. Diabetol. 14 (Jun. 2015) 70.
[15] M. Ohara, et al., Relationship between daily and day-to-day glycemic variability and increased oxidative stress in type 2 diabetes, Diabetes Res. Clin. Pract. 122 (Dec. 2016) 62–70.
[16] X. Tang, et al., Glycemic variability evaluated by continuous glucose monitoring system is associated with the 10-y cardiovascular risk of diabetic patients with well-controlled HbA1c, Clin. Chim. Acta 461 (Oct. 2016) 146–150.
[17] R.M. Bergenstal, Glycemic variability and diabetes complications: does it Matter? Simply put, there are better glycemic markers!, Diabetes Care 38 (8) (Aug. 2015) 1615–1621.
[18] B. Hirsch, Glycemic variability and diabetes complications: does it Matter? Of course it does!, Diabetes Care 38 (8) (Aug. 2015) 1610–1614.
[19] C. Fabris, et al., Glucose variability indices in type 1 diabetes: parsimonious set of indices revealed by sparse principal component analysis, Diabetes Technol. Therapeut. 16 (10) (Oct. 2014) 644–652.
[20] C. Fabris, A. Facchinetti, G. Fico, F. Sambo, M.T. Arredondo, C. Cobelli, Parsimonious description of glucose variability in type 2 diabetes by sparse principal component analysis, J. Diabetes Sci. Technol. 10 (1) (Jan. 2016) 119–124.
[21] G. Acciaroli, et al., Diabetes and prediabetes classification using glycemic variability indices from continuous glucose monitoring data, J. Diabetes Sci. Technol. 12 (1) (Jun. 2017) 105–113.
[22] "mosaicproject.eu." [Online]. Available: http://www.mosaicproject.eu/. (Accessed: 29 August 2017).
[23] American Diabetes Association, Diagnosis and classification of diabetes mellitus, Diabetes Care 37 (S1) (Jan. 2014) S81–S90.
[24] B. Kovatchev, C. Cobelli, Glucose variability: timing, risk analysis, and relationship to hypoglycemia in diabetes, Diabetes Care 39 (4) (Apr. 2016) 502–510.
[25] J.M. Wojcicki, 'J'-index. A new proposition of the assessment of current glucose control in diabetic patients, Horm. Metab. Res. 27 (1) (Jan. 1995) 41–42.
[26] D. Rodbard, New and improved methods to characterize glycemic variability using continuous glucose monitoring, Diabetes Technol. Therapeut. 11 (9) (Sep. 2009) 551–565.
[27] C.M. McDonnell, S.M. Donath, S.I. Vidmar, G.A. Werther, F.J. Cameron, A novel approach to continuous glucose analysis utilizing glycemic variation, Diabetes Technol. Therapeut. 7 (2) (Apr. 2005) 253–263.
[28] H. El-Laboudi, I.F. Godsland, D.G. Johnston, N.S. Oliver, Measures of glycemic variability in type 1 diabetes and the effect of real-time continuous glucose monitoring, Diabetes Technol. Therapeut. 18 (12) (Dec. 2016) 806–812.
[29] F.J. Service, G.D. Molnar, J.W. Rosevear, E. Ackerman, L.C. Gatewood, W.F. Taylor, Mean amplitude of glycemic excursions, a measure of diabetic instability, Diabetes 19 (9) (Sep. 1970) 644–655.
[30] P.A. Baghurst, Calculating the mean amplitude of glycemic excursion from continuous glucose monitoring data: an automated algorithm, Diabetes Technol. Therapeut. 13 (3) (Feb. 2011) 296–302.
[31] M. Wiley, R. Bunescu, C. Marling, J. Shubrook, F. Schwartz, Automatic detection of excessive glycemic variability for diabetes management, in: Presented at the 2011 10th International Conference on Machine Learning and Applications and Workshops (ICMLA), Honolulu, HI, vol. 2, 2011, pp. 148–154.
[32] C.R. Marling, J.H. Shubrook, S.J. Vernier, M.T. Wiley, F.L. Schwartz, Characterizing blood glucose variability using new metrics with continuous glucose monitoring data, J. Diabetes Sci. Technol. 5 (4) (Jul. 2011) 871–878.
[33] C. Marling, M. Wiley, R. Bunescu, J. Shubrook, F. Schwartz, Emerging applications for intelligent diabetes management, AI Mag. 33 (2) (Jun. 2012) 67–78.
[34] J. Schlichtkrull, O. Munck, M. Jersild, The M-value, an index of blood-sugar control in diabetics, Acta Med. Scand. 177 (1) (Jan. 1965) 95–102.
[35] N.R. Hill, P.C. Hindmarsh, R.J. Stevens, I.M. Stratton, J.C. Levy, D.R. Matthews, A method for assessing quality of control from glucose profiles, Diabet. Med. 24 (7) (Jul. 2007) 753–758.
[36] D. Rodbard, Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control, Diabetes Technol. Therapeut. 11 (S1) (May 2009). S-55–S-67.
[37] B.P. Kovatchev, D.J. Cox, L.A. Gonder-Frederick, W. Clarke, Symmetrization of the blood glucose measurement scale and its applications, Diabetes Care 20 (11) (Nov. 1997) 1655–1658.
[38] M. Reddy, et al., Glycemic variability and its impact on quality of life in adults with type 1 diabetes, J. Diabetes Sci. Technol. 10 (1) (Aug. 2015) 60–66.
[39] M.-K. Hu, Visual pattern recognition by moment invariants, IEEE Trans. Inf. Theor. 8 (2) (Feb. 1962) 179–187.
[40] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag New York, New York, NY, 2009.
[41] P. Augstein, et al., Q-Score: development of a new metric for continuous glucose monitoring that enables stratification of antihyperglycaemic therapies, BMC Endocr. Disord. 15 (22) (May 2015).
[42] B. Hirsch, A.K. Balo, K. Sayer, A. Garcia, B.A. Buckingham, T.A. Peyser, A simple composite metric for the assessment of glycemic status from continuous glucose monitoring data: implications for clinical practice and the artificial pancreas, Diabetes Technol. Therapeut. 19 (S3) (Jun. 2017). S-38–S-48.
[43] K.J. Ruedy, C.G. Parkin, T.D. Riddlesworth, C. Graham, Continuous glucose monitoring in older adults with type 1 and type 2 diabetes using multiple daily injections of insulin: results from the DIAMOND trial, J. Diabetes Sci. Technol. 11 (6) (Nov. 2017) 1138–1146.
[44] C.L. Chan, et al., Alternate glycemic markers reflect glycemic variability in continuous glucose monitoring in youth with prediabetes and type 2 diabetes, Pediatr. Diabetes 18 (7) (Nov. 2017) 629–636.
[45] F. Li, et al., Features of glycemic variations in drug naïve type 2 diabetic patients with different HbA 1c values, Sci. Rep. 7 (1) (May 2017) 1583.